

Molecular Biology and Biochemistry of Differentiation

by Stephen E. Harris*

Recent work in the field of molecular biology and differentiation has been directed towards an assessment of the number of different genes involved in the development and differentiation process. By the techniques of RNA-DNA hybridization to single copy DNA, it appears that some 60,000-200,000 different RNA sequences are expressed during embryonic development in the mouse. The differentiating brain shows the highest degree RNA transcription diversity. The new technique of cDNA-poly(A)-containing RNA hybridization is described. The nuclear poly(A)-containing RNA appears to reflect the high complexity of sequences as determined in RNA-DNA experiments described above. The cytoplasmic poly(A)-containing RNA or messenger RNA appears to represent a subset of approximately 10-20% of the information in the nuclear poly(A)-RNA. Also, the major portion or frequency class of messenger RNA in different organs such as kidney, spleen, and liver are common to these different organs. However, brain, although containing many of the cytoplasmic messenger RNA sequences found in liver, kidney, and spleen, has a large class of messenger RNA sequences which are specific to the nervous system.

At the heart of understanding the mechanisms involved in cellular differentiation in eukaryotic organism is an assessment of the genomic output, both qualitative and quantitative, during embryonic development and in the final differentiated organ systems. Since each somatic cell contains the same genome in terms of DNA content and sequence information, the mechanisms underlying cellular differentiation must be reflected in differential gene activation and repression of specific DNA sequences at various times during development. There must be control mechanisms capable of activating and selecting part of the total genetic potential while repressing other regions at different developmental stages or in different cell types. It is assumed that specific cellular products produced by one cell can and do influence the fate of gene expression patterns of other cell types in the surrounding environment of the differentiating system. There must be

many genes, both structural and regulatory, involved in cellular processes of differentiating cells as well as in the final differentiated adult organ. We assume that many of these expressed genes will be both common to different cell types as well as genes which will be specific and characteristic of a given cell or organ.

Recent work in the field of the molecular biology of differentiation in eukaryotes, and in particular in the mouse system, has concerned itself with analysis of the informational content of RNA transcripts during development. With this information one can obtain a feel for the number of potential genes that are involved in the stabilization of the final differentiated state. The techniques of nucleic acid hybridization have provided us with useful information about genomic output in normal and abnormal development, and this tool could in the future be very useful for assessing the effect of various toxic and teratogenic compounds on various organ systems during embryonic development.

Although much information has accumulated on biochemical and genetic aspects of the pre-

*Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030.

implantation mouse and rabbit embryo (1), this brief review will be more restricted to an analysis of genomic output at later stages of development and in final differentiated organs. The techniques of nucleic acid hybridization will be central to answering the following questions. First, what is the expressed genomic output during embryonic development in the mouse, (that is, the total informational content in the RNA transcripts or number of potential genes expressed during development)? For convenience, we define an average gene as a 1000 nucleotides of information. This is enough information to code for a protein with a molecular weight of 38,000. The second question we will be concerned with is the total potential genes expressed in RNA in various differentiated organs of the mouse. How much of this genomic output is common and how much is specific for the various organs looked at? Finally, we can ask how much of the potential genetic information is actually transported to the cytoplasm and translated into protein (structural genes) and how much is either destroyed or processed and remains in the nucleus of the cell? With this information we may obtain an idea of at what level does the greatest organ specificity reside in terms of these RNA transcripts (nuclear vs. cytoplasm).

First, we must consider something about DNA sequence content and organization in the mouse. The DNA of most higher organisms, in particular the mouse, is said to contain basically three classes of DNA based on differences in their reannealing kinetics (Fig. 1). The highly repetitive DNA, the satellite DNA, contains over a million copies of a simple repeating sequence. The middle repetitive DNA, which in the mouse genome are present from a few hundred to several thousand copies, represents about 25% of the total mass of DNA in a typical mouse cell. The major class of DNA (60%) reanneals with hybridization kinetics indicating only one copy per haploid genome. It is this unique DNA or single copy DNA which we will mainly be concerned about since RNA hybridization data with this DNA fraction can yield quantitative information about total informational output for a particular cell, tissue, or organ (1).

Sequence Organization and Primary Transcript

Recent evidence has indicated that a major fraction of the middle repetitive DNA sequences are short [200-400 nucleotide pairs (NTP)] and are intimately interspersed with unique DNA se-

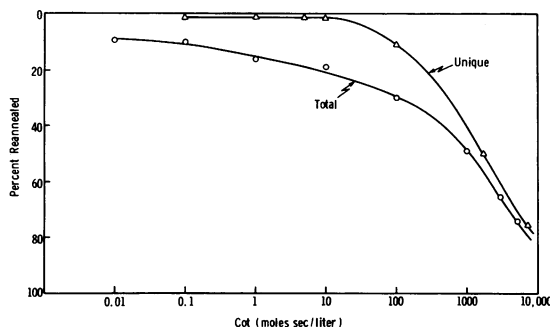


FIGURE 1. Total mouse DNA was sheared to approximately 350 nucleotides and allowed to reanneal at 0°C in 0.18M Na⁺ or the equivalent salt. At various *Cot* values (moles of nucleotides/liter × time), samples were removed and analyzed by hydroxylapatite. The unique DNA was isolated by first reannealing total mouse DNA to a *Cot* value of 500 and then isolating the DNA which does not bind hydroxylapatite at 60°C in 0.14M phosphate buffer.

quences 1000-2000 NTP long (2). The primary transcript of DNA or heterogeneous nuclear RNA (HnRNA) also contain short repetitive elements interspersed with unique DNA sequences (3). Whether the large HnRNA (5,000-20,000 nucleotides) serve as precursor to the smaller messenger RNA in the cytoplasm is still open to question, but some fraction of nuclear RNA must serve as messenger RNA (4). The major portion of repetitive transcripts remain in the nucleus or are broken down (5).

Expressed Genomic Output or Transcriptional Complexity

Once having isolated the unique DNA of the mouse by hydroxylapatite chromatography and establishing that the repetitive DNA has been removed (Fig. 1) and that the analytical complexity or total unique genome complexity of the DNA is 2×10^9 NTP, one may now determine the expressed genomic output or transcriptional complexity of an RNA isolated from a particular cell, tissue, or organ. For example, as in Figure 2, a given RNA hybridized with 2% of the single copy DNA. This would then represent an RNA complexity of 4×10^7 NT or 40,000 different genes, each a 1000 nucleotides in length.

Transcriptional Complexity during Development In the Mouse

Recently, Church and Schulz (1) have looked at the transcriptional complexity of RNA isolated

from various embryonic stages of mouse development by hybridization to radioactive unique DNA. The results are summarized in Figure 3. From the 8-16 cell stage where there are some 10,000 potential genes of 1,000 NT each to the blastocyst, involves approximately a 3-fold increase in transcriptional complexity (30,000 potential genes). From the blastocyst to term there is another 7-fold increase in transcriptional complexity (ca. 200,000 potential genes). This high RNA complexity presumably reflects the differentiation of the various organs and tissues. As we will see later, much of it is contributed by the nervous system of the mouse.

I should point out the reason one refers to the transcriptional complexity as potential genes is because it is not clear what the function is of much of this expressed genetic information. We do know that some fraction of the potential gene transcripts will be transported to the cytoplasm and become associated with polysomes. This information must represent the activity of the structural genes (messenger mRNA), since presumably this RNA information will be translated into proteins. Suffice it to say much of the transcriptional complexity of a given cell never reaches the cytoplasm and may play other vital roles in the cellular machinery.

Total Genomic Output in Different Adult Organs and in Different Organs during Development

In Figure 4 is shown an example of data obtained from similar RNA excess hybridization ex-

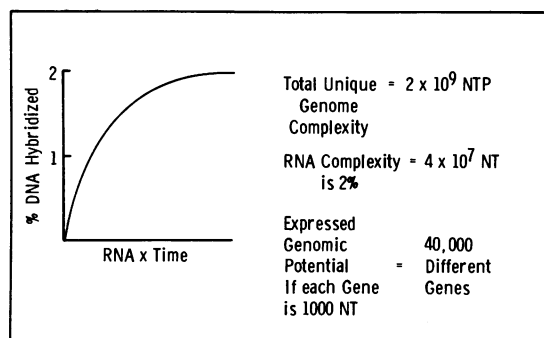


FIGURE 2. Method for measuring expressed genomic output or total transcriptional complexity of an organ. Total RNA is allowed to hybridize to radioactive unique DNA in an RNA driven manner. Usually RNA at 10-20 mg/ml is hybridized with a few micrograms/ml of unique DNA for several days. The samples are then analyzed for RNA-DNA and DNA-DNA hybrids after no treatment and extensive RNase treatment, respectively. The amount of DNA in true RNA hybrid is then calculated.

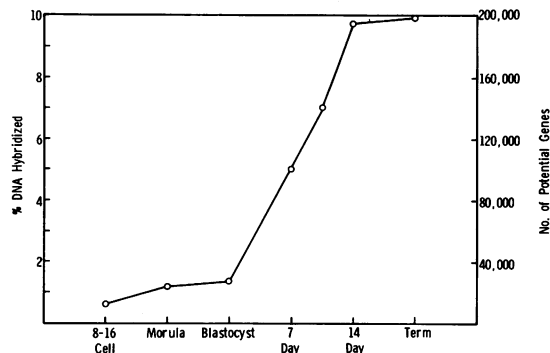


FIGURE 3. Expressed genomic output during mouse developments. Experiments done essentially as described in Figure 2. Data of Church and Schultz (1).

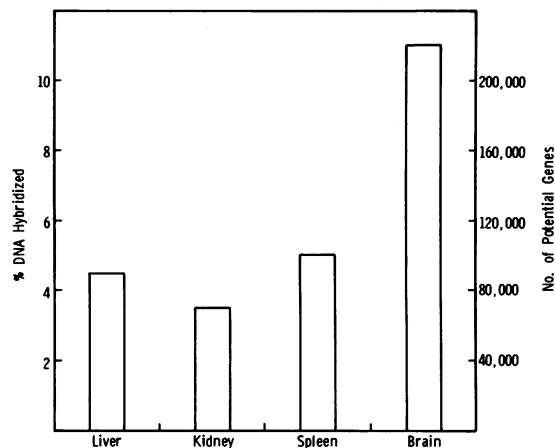


FIGURE 4. Total RNA from the various organs was again hybridized to small amounts of ^3H -unique mouse DNA and the amount of RNA-DNA hybrid determined. Data of Grouse et al. (6).

periments with mouse unique DNA (6). Liver, kidney, and spleen all hybridize to about 3-4% or a genomic output which represents some 60,000 to 80,000 different potential genes. Brain, on the other hand, has a total genomic output 3 to 4 times that of liver, kidney, or spleen. Presumably, this reflects the wide variety of cell types found in the brain tissue. Interestingly, as shown in Figure 5, the brain genomic output increases during late embryonic development and postnatally while the liver actually shows a slight decrease in the number of potential genes expressed postnatally. The effect of various toxic elements such as mercury chloride on the development and differentiation of the brain both during *in utero* life as well as after birth might be an area to analyze in terms of changing genomic output.

Overlapping Genomic Output in Different Mouse Tissues

An obvious question about the genomic output in different organs is how different or similar are the total RNA transcripts from these organs? This is a somewhat difficult question but can be approached by an additive RNA saturation experiment; that is, the RNA from liver and spleen are mixed in equal proportions and hybridized to the unique DNA (6). As shown in Figure 6, if all the sequences were different in liver and kidney, then the additive hybridization would be about 8% or 160,000 potential expressed genes. Experimentally, a 7% hybridization value is obtained. Therefore, approximately 20,000 sequences are common to both tissues. Similar experiments as shown in Figure 6, indicate that liver and spleen share some 60,000 different RNA sequences. It should be pointed out that this sort of additive hybridization experiment does not give one information about specific sequences to a given tissue—only the sequences that are common between those two organs and this is probably only a minimal estimate.

Polyadenylic-Containing RNA: Cytoplasmic Messenger RNA and Nuclear RNA

Another approach to the tissue and organ specificity and overlap of RNA transcripts may be looked at by another recent technique. First, since discovering that a large fraction of cytoplasmic messenger RNA and a relatively large proportion of nuclear RNA have polyadenylic acid (poly A) tails on the 3'-end, a new approach has opened up to determining the genomic information of complexity of these poly A-RNAs in different organs (7).

In Figure 7 is presented the scheme for the production and utilization of a complementary DNA probe to a given poly A-containing RNA population. With the use of a short oligodeoxythymidyl acid primer and the enzyme reverse transcriptase and ^3H -deoxynucleoside triphosphates as substrate, a highly specific and sensitive probe can be synthesized. By hybridizing the radioactive complementary DNA (cDNA) to the RNA from which it was synthesized, the complexity of that poly A-RNA population can be determined by the rate at which hybrids are formed with respect to time and RNA concentration (Cr_t). This principle is demonstrated in Figure 8. If a ^3H -com-

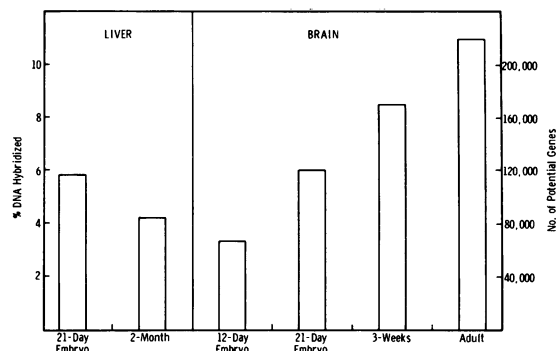


FIGURE 5. Developmental changes in liver and brain expressed in unique DNA information. Data of Grouse et al. (6).

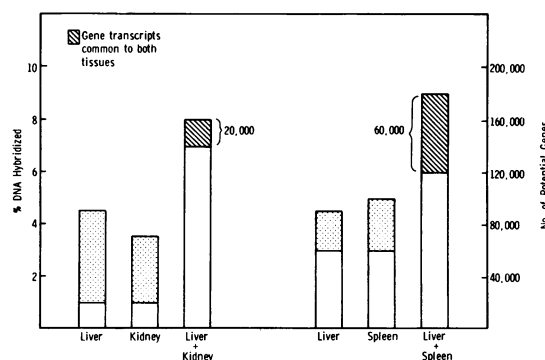


FIGURE 6. Common unique DNA transcripts in liver, spleen, and kidney. Data of Grouse et al. (6).

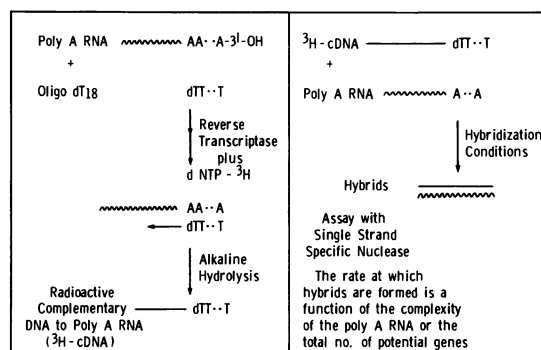


FIGURE 7. Synthesis of complementary DNA to poly A-containing RNA and use as a probe to estimate transcriptional complexity. The principal assumption in this technique is that the enzyme, reverse transcriptase synthesizes the cDNA sequences in the same proportion as found in the poly A-containing RNA.

plementary DNA is made to a pure messenger RNA, such as ovalbumin messenger RNA, the rate at which a sample of pure mRNA_{ov} will react or back hybridize to the radioactive cDNA_{ov} will be extremely fast since the rate is determined by the complexity, and one gene product obviously has a low complexity compared to total RNA with thousands of different sequences. On the other hand, if a poly A containing RNA population contains a 1000 different sequences, each 1000 nucleotides long, then a complementary DNA preparation from that RNA would hybridize back to the RNA at a rate a thousand times slower. Naturally, if one were to mix a pure messenger RNA and poly A containing RNA having 1000 different sequences in a one-to-one proportion, make the cDNA to this mix, and then obtain the back hybrid, the kinetic analysis or "Crot curve" would show two components with first-order rate constants varying by a factor of 1000 (Fig. 8). This, of course, is an ideal situation, and in reality when one makes cDNA to poly A containing RNA from a cytoplasmic or nuclear fraction and back hybridizes the respective RNA to its cDNA, complex kinetic curves are generated which are best resolved by computer into various first-order kinetic classes or frequencies of sequences. In the idealized example (Fig. 8), there are two frequency classes, one with 1 sequence and a second with 1000 sequences, each representing 50% of the mass.

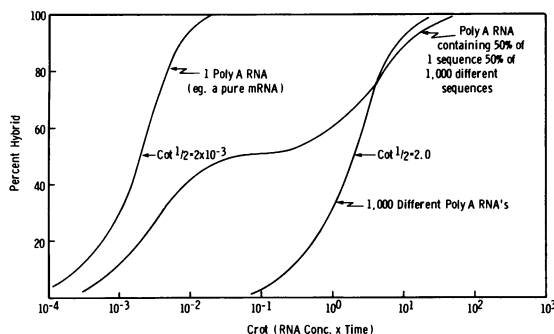


FIGURE 8. Kinetic analysis of poly A-containing RNA by use of ³H-complementary DNA. This is a simple example illustrating the principle of analyzing a complex mixture of different sequences in different concentrations in a given RNA preparation.

Comparison of Nuclear RNA and Messenger RNA

Previous results in embryonic systems as well as normal or transformed cells have indicated

that much of the labeled nuclear RNA of HnRNA is broken down to smaller RNAs as well as down to nucleotides, and only a small portion (5-10%) is transferred to the cytoplasm with time (4,5). Recent experiments by Davidson and Britten's group working with sea urchin embryos have clearly shown that the complexity of the nuclear RNAs which hybridize to unique DNA is some ten times that of the messenger RNA isolated from the same group of embryonic cells (8). One may now ask how much of the nuclear poly A-containing RNA information is transferred to the cytoplasm; by cDNA-back hybrid analysis we may get an indication of the sequence complexities of these two RNA populations.

Paul's group, using mouse Friend cells, have recently analyzed the cytoplasmic and the nuclear poly A-containing RNA populations (9) by cDNA analysis. Their results can be summarized as follows (Fig. 9). The hybridization curves for the poly A messenger RNA are complex and can be resolved into at least three frequency classes or components. The average frequency class would indicate that approximately 4,000 to 20,000 different sequences may actually be translated in the Friend cell. On the other hand, when one analyzes the nuclear cDNA-nuclear poly A-RNA hybridization curve, which again can be divided into at least two frequency classes, it has a major first-order component ($Crot_{1/2} = 400$) indicating 80,000 potential genes. As you recall, this number is reasonably close to the value arrived at by the independent method of RNA saturation to unique DNA for various tissues like spleen, kidney, and liver (60,000-100,000 potential genes each 1000 nucleotides long). Suffice it to say, the total poly A-RNA complexity or nuclear poly A-RNA complexity tends to reflect the high complexity found

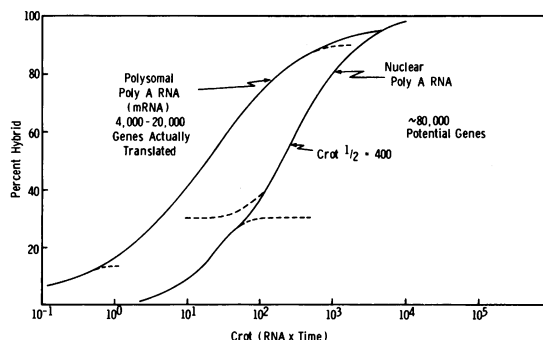


FIGURE 9. Comparison of nuclear poly A-containing RNA transcription complexity and polysomal Poly A-containing RNA (mRNA) by use of ³H-complementary DNA probes in mouse Friend cells. Adapted from Getz et al. (9).

in total RNA hybridized to unique DNA, while the cytoplasmic or messenger poly A-containing RNA appears to be a subset of those nuclear sequences. How the information in nuclear RNA is processed and selected to result in the subset of sequences which ultimately are translated could be an important point of control in both normal and abnormal development and differentiation.

Overlapping Poly A-Messenger RNA Populations In Different Tissues and Organs

The complementary DNA-poly A-containing RNA hybridization method offers an opportunity to assess the overlapping and tissue specific poly A-containing RNA sequences. In fact, the possibility of purifying a cDNA probe specific to a given specific state of an organ should be possible by obtaining various heterologous cDNA-poly A-containing RNA hybridizations and isolating the cDNA that does not enter into hybrid. For example, Ryffel and McCarthy (10) have made a cDNA to total cytoplasmic messenger poly A-containing RNA from mouse brain, hybridized the cDNA with liver, spleen, and kidney messenger RNA (60% hybrid), and then isolated by hydroxyapatite chromatography the cDNA that did not form hybrids (~40%). This cDNA is relatively specific for brain tissue.

In a similar matter, the overlap of messenger RNA populations from different tissues was studied by this method. A complementary DNA probe was made to L-cell cytoplasmic poly A messenger RNA. The kinetics of hybridization of the cDNA and the RNA from which it was made were analyzed by computer (Fig. 10). The curve can be resolved into three components, the third or high complexity component representing some 13,000 structural genes (330 amino acids each).

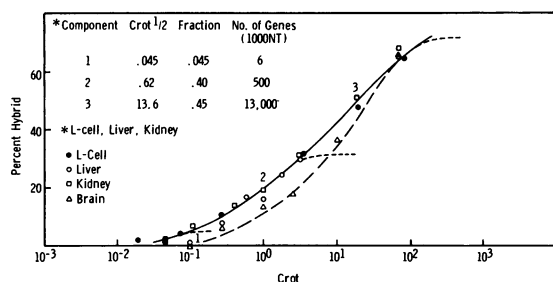


FIGURE 10. Hybridization of complementary DNA synthesized from cytoplasmic poly A-RNA from L-cells with cytoplasmic poly A-RNA (mRNA) from different mouse organs. Adapted from Ryffel and McCarthy (10).

What is of interest here is the fact that liver, kidney, spleen, and brain cytoplasmic poly A messenger RNA can all hybridize and protect most of the L-cell cDNA. This indicates that these tissues and cells all have in common a large fraction of their messenger RNA population. However, it should be pointed out, that the RNA curve with brain RNA is shifted to higher Crot value due to the large fraction of the RNA which is specific to that tissue. It should also be pointed out that a 1 or 2% difference in the hybridization values of the different tissues could represent 100 to 200 different organ- or tissue-specific messenger RNAs, and the technique as presently used is in no way sensitive enough to pick up differences of 1-2% (10).

In Figure 11, is summarized the information briefly reviewed in this paper. The main points are that the genome puts out a lot of information apparently not involved in synthesizing proteins and the major portion of the information that does eventually arrive in the cytoplasm is common to a variety of tissues and organs. This might indicate that the nuclear RNA will show the greatest tissue and organ diversity. This would in turn point to this high-complexity nuclear RNA as important in processing of the primary transcript or possibly having some role in the regulation of development and differentiation.

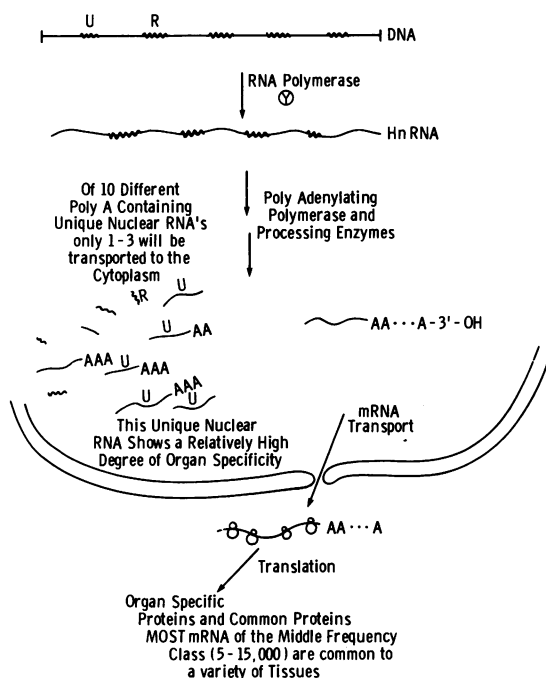


FIGURE 11. Summary: RNA informational flow from the genome to the cytoplasm.

REFERENCES

1. Church, R. B., and Schultz, G. A. Differential gene activity in the pre-and postimplantation mamalian embryo. In: *Current Topics in Development Biology*, Vol. 9, Academic Press, New York, 1974, p. 179.
2. Davidson, E. H., et al. Comparative aspects of DNA organization in metazoa, *Chromosoma* 51: 253 (1975).
3. Holmes, D. S., and Bonner, J. Interspersion of repetitive and single-copy sequences in nuclear ribonucleic acid of high molecular weight. *Proc. Natl. Acad. Sci. U.S.* 71:1108 (1974).
4. Darnell, J. E., Jr., The origin of mRNA and the structure of the mammalian chromosome. In: *The Harvey Lectures Series* 69, 1973-1974. Academic Press, New York, 1975, p. 1.
5. Lewin, B. Unit of transcription and translation: sequence components of heterogenous nuclear RNA and messenger RNA. *Cell* 4: 77 (1975).
6. Grouse, L., et al. Hybridization of ribonucleic acid with unique sequences of mouse deoxyribonucleic acid. *Biochemistry* 11: 798 (1972).
7. Bishop, J. O., et al. Three abundance classes in Hela cell messenger RNA. *Nature* 250: 199 (1974).
8. Hough, B. R., et al. Sequence complexity of heterogenous nuclear RNA in sea urchin embryos. *Cell* 5: 291 (1975).
9. Getz, M. J., et al. A kinetic estimation of base sequence complexity of nuclear poly (A)-containing RNA in mouse Friend cells. *Cell* 4: 121 (1975).
10. Ryffel, G. O., and McCarthy, B. J. Complexity of cytoplasmic RNA in different mouse tissues measured by hybridization of polyadenylated RNA to complementary DNA. *Biochemistry* 14: 1379 (1975).